Setting standards in language testing and assessment: past, present and future: the impact of the CEFR

Neus Figueras
Generalitat de Catalunya, Spain
University of Barcelona

5th British Council New Directions in Language Assessment Conference – Post Symposia
April 2018

Outline

- Some terms → **WHAT**

- Quality in language learning, teaching and assessment

   instruments

   decisions   **WHY**

- Transparency and Accountability

- The CEFR → **HOW**

What is a standard?

- A principle of honesty and integrity.
- An authoritative exemplar of perfection.

- A definite level of excellence or attainment.
- A recognised degree of proficiency.

- The measure of what is adequate for a purpose.
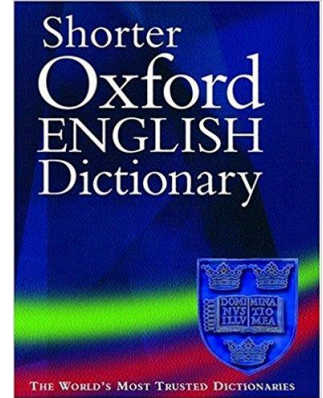- A norm of comparison or judgement.
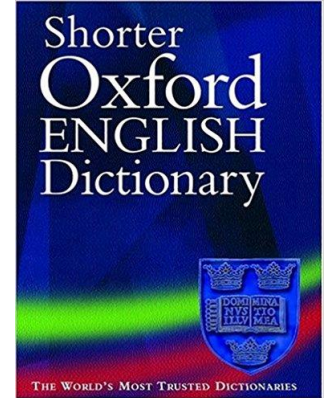
What is a standard?

- A principle of honesty and integrity.
- An authoritative exemplar of perfection.

- A definite level of excellence or attainment.
- A recognised degree of proficiency.

- The measure of what is adequate for a purpose.
- A norm of comparison or judgement.

- Beginner
- Fase Beginner
- Elementary
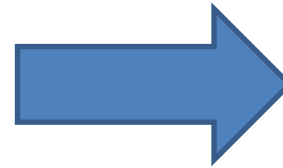- Early-Intermediate
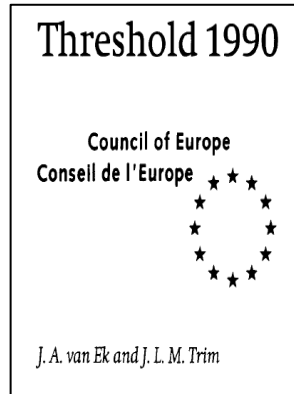- Intermediate
- Upper Intermediate
- Advanced
- …/…

Exam "X"

The present

- CEFR
- ACTFL
- ACTFL NCSSFL can dos
- Canadian Benchmarks
- ASLPR - Australian Second Language Proficiency Ratings
- CEFR – J
- CSA

../..

The CEFR in history

The Major Project in Modern Languages
Council of Europe 1964

Threshold 1990

Council of Europe
Conseil de l'Europe

J. A. van Ek and J. L. M. Trim

The Ruschlikhon Symposium 1991
Transparency and coherence in language learning in Europe

Common European Framework
of Reference for Languages:
Learning, teaching, assessment

Modern Languages Division, Strasbourg
CAMBRIDGE UNIVERSITY PRESS

January 2009

Relating Language Examinations to the CEFR

Language Policy Division

www.coe.int/lang

CEFR Companion Volume with extended illustrative descriptors

Education Department

September 2017
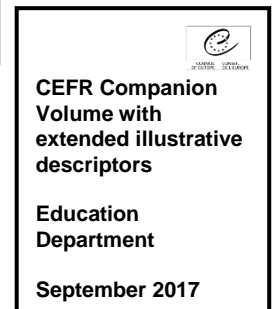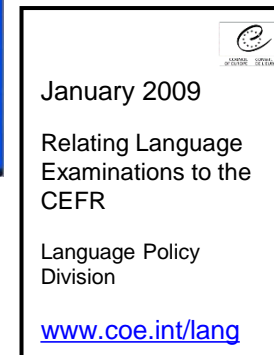
Impact

- Translated into 40 languages.
- Used in curriculum design, textbooks, ... (beyond Europe).

  Common currency
- Used in language testing and assessment.
- Used in international projects (Dialang, Surveylang,...).
- Triggered research and discussion.

Although..
- Not yet used to its full potential.

Issues and criticisms

- Adoption → Adaptation
- PLD (Performance level descriptors) → PLL (Performance level labels)
- Reification

- Practitioners - not accesible/user-friendly, imprecise
              - descriptors -> objectives
- SLA - language proficiency re. language development
       - language proficiency re. cognitive abilities
- Testers - missing/unbalanced scales & descriptors
           - underspecification

Reactions of the Council of Europe

- Recommendations -> Policy makers

- Toolkit -> Practitioners

- Companion Volume -> Wide audience (?)

Recommendation CM/Rec (2008/7)

Of the Committee of Ministers to member states on the use of the Council of Europe's CEFR and the promotion of Plurilingualism *(Adopted by the Committee of Ministers on 2 July 2008)*

General principles and measures to be implemented by authorities responsible for language education at national, regional and local level.

Specific measures aimed at policy making, curriculum and textbook development, teacher training and assessment.

Appendix

Explanatory Notes

The CEFR toolkit and resources

www.coe.int

The learner and the languages present in school

Regional, minority and migration languages

LANGUAGE(S) OF SCHOOLING

Foreign languages – modern and classical

Language as a subject

Language(s) in other subjects

The Manual offers guidance to users on how to

• Describe their examination coverage, administration and analysis procedures;

• Relate results reported from the examination to the CEFR levels presented in Chapter 3 in the CEFR;

• Provide supporting evidence reporting the procedures followed to do the above.

(Manual 2009: 2)

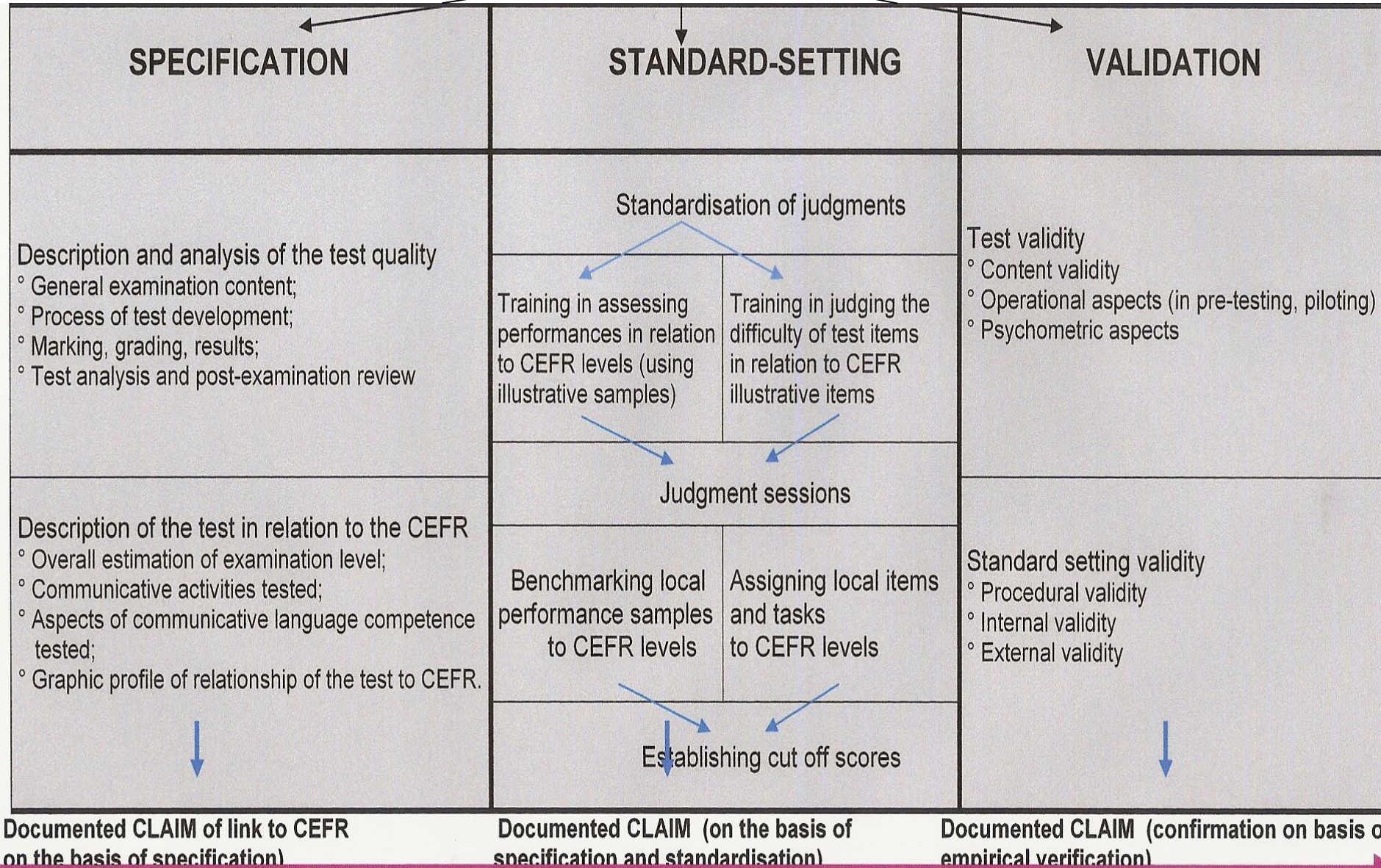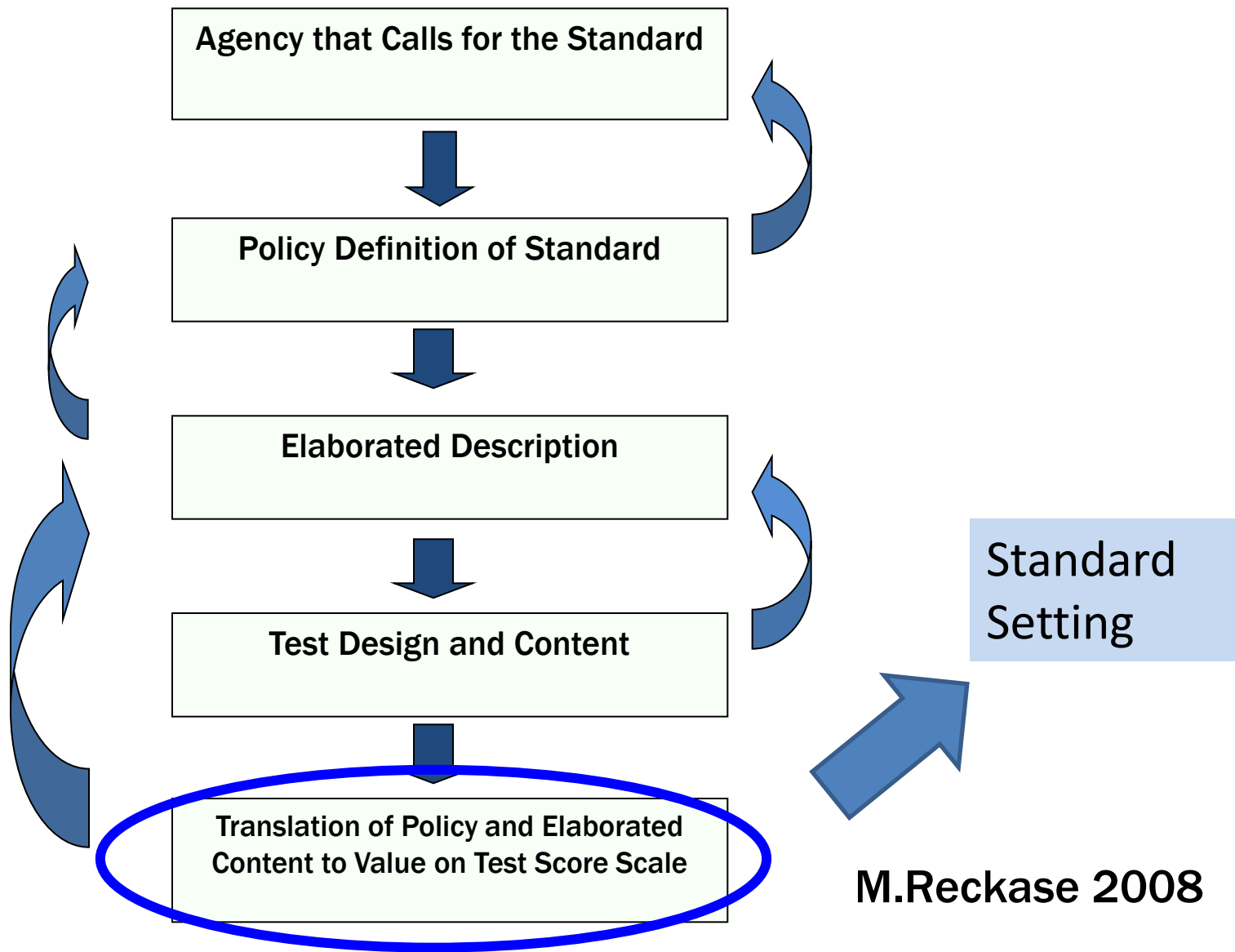# Building an argument

Familiarisation with the CEFR

SPECIFICATION	STANDARD-SETTING	VALIDATION
Description and analysis of the test quality ° General examination content; ° Process of test development; ° Marking, grading, results; ° Test analysis and post-examination review	Standardisation of judgments Training in assessing performances in relation to CEFR levels (using illustrative samples) | Training in judging the difficulty of test items in relation to CEFR illustrative items Judgment sessions	Test validity ° Content validity ° Operational aspects (in pre-testing, piloting) ° Psychometric aspects
Description of the test in relation to the CEFR ° Overall estimation of examination level; ° Communicative activities tested; ° Aspects of communicative language competence tested; ° Graphic profile of relationship of the test to CEFR.	Benchmarking local performance samples to CEFR levels | Assigning local items and tasks to CEFR levels Establishing cut off scores	Standard setting validity ° Procedural validity ° Internal validity ° External validity
Documented CLAIM of link to CEFR on the basis of specification)	Documented CLAIM (on the basis of specification and standardisation)	Documented CLAIM (confirmation on basis o empirical verification)

Agency that Calls for the Standard

Policy Definition of Standard

Elaborated Description

Test Design and Content

Translation of Policy and Elaborated Content to Value on Test Score Scale

Standard Setting

M.Reckase 2008

What is standard setting?

Standard setting is a general label for a number of approaches commonly used to identify test scores that support decisions about test takers' (candidates') level of knowledge, skill, proficiency, mastery, or readiness.

Standard setting, according to Cizek (1993), is "the proper following of a prescribed, rational system of rules or procedures resulting in the assignment of a number to differentiate between two or more states or degrees of performance" (p. 100).

Fundamentally, standard setting involves the development of a policy about what is required for each level of performance. This policy is stated in the performance standards and implemented through the cutscores.

Cizek and Bunch (2007) summarized the judgmental nature of standard setting thus:

"To some degree, then, because standard setting necessarily involves human opinions and values, it can also be viewed as a nexus of technical, psychometric methods and policy making"

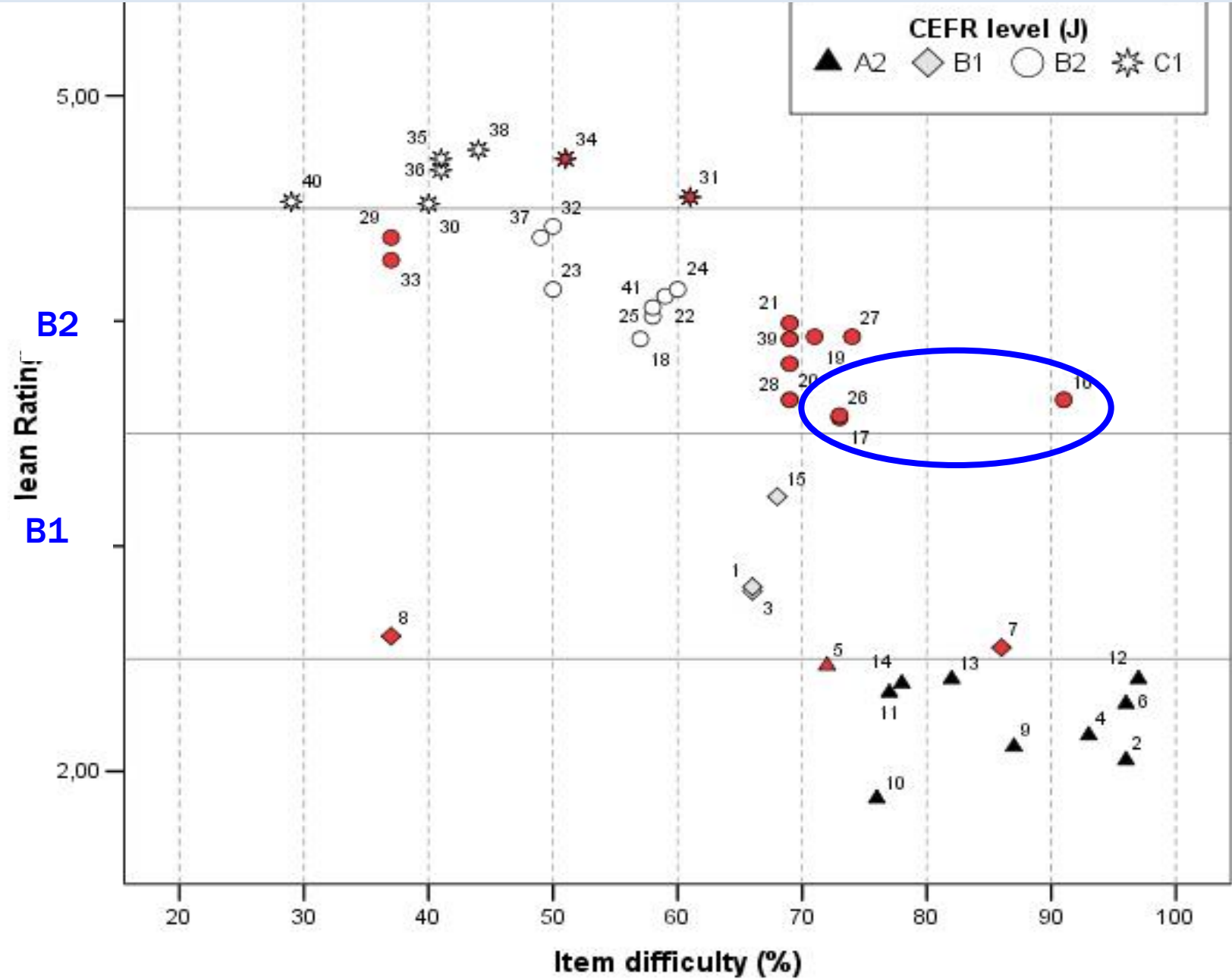Tannenbaum and Wylie (2008:2)

How do test content and performance level descriptors match?

How does difficulty relate to performance levels?

What should inform decisions on cut scores?

How do I know my B1 is your B1?

Is my standard your standard?

What's your B1 like?

What's your standard like?

"It is imperative that examinations today are

- ☐ Honest
- ☐ Reliable
- ☐ Valid
- ☐ Transparent
- ☐ Portable "

John Trim's address at the ACTFL-CEFR Conference in Provo, USA.

August 4th, 2011

D**. LINKAGE TO THE COMMON EUROPEAN FRAMEWORK – CEFR**

1. Have issues of test quality and impact been fully addressed? (see EALTA Guidelines Section C). Is information on the quality of the test publicly available?

2. Are those involved in test development all familiar with the CEFR? Have familiarisation with the CEFR procedures been followed? What evidence is there of the procedures followed and the corresponding results?

3. Have the test content and the test specifications been analysed in relation to the CEFR descriptors? What evidence is there of the procedures followed and the corresponding results?

4. Have standardisation procedures been completed for written and spoken performances? Which ones? By whom? What evidence is there of the procedures followed and the corresponding results?

5. Are local performance samples benchmarked to the CEFR publicly available?

6. Have standardisation procedures been completed for test items targeting receptive skills? Which ones? By whom? What evidence is there of the procedures followed and the corresponding results?

7. Are local CEFR-linked items and tasks publicly available?

8. What standard setting procedures have been used to establish cut-off scores for the relevant CEFR level(s)? How many judges have been involved? Is there a standard setting report publicly available?

9. Has validity evidence been collected covering the process of linkage? Is it publicly available?

10. Is there a scheme of level setting that guarantees quality standards and linkage to the CEFR? Is it publicly available?

- Recommendations -> Policy makers

- Toolkit -> Practitioners

- Companion Volume -> Wide audience (?)

Introduction

Changes to descriptors + additions

New scales (34)
 Mediation
 Plurilingualism/Pluriculturalism
 Online interaction
 Reading as a leisure activity

Appendices

"It is important to note that the additions do not impact on the construct described in the CEFR, or on its Common Reference Levels. The Companion Volume, and in particular the descriptors for new areas, represent an enrichment of the original descriptive apparatus." CV 2017:22

Work will continue….

- How can a real and stronger impact of the CEFR in the classroom be achieved? Will the CEFR Companion Volume facilitate it?

- How will the changes, additions and new scales impact Teaching. Learning and Assessment?

- What difficulties can be predicted?

- What can be done to make the CEFR CV more accesible to practitioners?

References

- Figueras, N. and Noijons, J. (Eds) (2009). *Linking to the CEFR levels, Research Perspectives.* CITO, Council of Europe and EALTA.   www.ealta.eu.org/resources. html
- Figueras, N.; Kaftandjieva, F.; Takala, S (2013) Relating a Reading Comprehension Test to the CEFR levels. A case of standard setting in practice. Canadian Modern Language Review.
- Harsh, C., Pant, H.A., Köller, O. (Eds) (2010) *Calibrating Standards-based Asssessment Tasks for English as a First Foreign Language – Standard-setting Procedures in Germany.* Waxmann Verlag GmbH. Munster
- Hawkins, J.A. and Dilipovic, L. (2012) *Criterial features in L2 English: Specifying the reference levels of the CEFR. English Profile Studies 1.*   Cambridge, UK. Cambridge University Press.
- Kaftandjieva, F. (2010) Methods for setting cut scores in criterion-referenced achievement tests: A comparative analysis of six methods with an application to tests of reading in EFL . Arnhem, The Netherlands: CITO. www.ealta.eu.org/resources.
- Livingston, S. and Zieky, M. (2006) *A Primer on SettinRTg Cut Scores on Tests of Educational Achievement. ETS* https://www.ets.org/research/policy_research_reports/cut-scores
- Martyniuk, W. (Ed) (2010) *Aligning tests with the CEFR. Reflections on using the Council of Europe draft Manual.* Cambridge UK. Cambridge University Press.
- Tannenbaum, R. J., & Wylie, E. C. (2008) Linking English-language test scores onto the Common European Framework of Reference: An application of standard setting methodology (TOEFL iBT Series Report No. 06). Princeton, NJ: Educational Testing Service http://www.ets.org/research/policy_research_reports/publications/report/2008/hspt